Submitted to Statistical Science

Qualitative Robust Bayesianism and the Likelihood Principle

Conor Mayo-Wilson and Aditya Saraf

The central question of this paper is, "Which controversial statistical principles (e.g., likelihood, conditionality, etc.) are special cases of principles for good scientific reasoning?" We focus on the likelihood principle [Birnbaum, 1962, Berger and Wolpert, 1988], the weak law of likelihood [Sober, 2008], and the strong law of likelihood [Edwards, 1984, Hacking, 1965, Royall, 1997], as these principles are critical for foundational debates in statistics.

Our central question is motivated by the following observation. Despite the ubiquity of statistics in contemporary science, many successful instances of scientific reasoning require only comparative, non-numerical judgments of evidential strength. For instance, when Lavoisier judged that his data provided good evidence against the phlogiston theory of combustion and for the existence of what we now call "oxygen", no statistics was invoked. Statistical reasoning, therefore, is just one type of scientific reasoning, and so norms for statistical reasoning should be special cases of norms for good scientific reasoning.

Our thesis is that likelihood principle (LP) and *weak* law of likelihood (LL) generalize naturally to settings in which experimenters are justified only in making comparative, non-numerical judgments of the form "A given B is more likely than C given D." Specifically, our main results show that, just as LP characterizes when all Bayesians (regardless of prior) agree that two pieces of evidence are equivalent, a qualitative/non-numerical version of LP provides sufficient conditions for agreement among experimenters' whose degrees of belief satisfy only very weak "coherence" constraints (i.e., ones that do not entail the probability axioms). We prove a similar result for LL. In contrast, the *strong* law of likelihood (LL$^+$) – which asserts that the likelihood ratio is a measure of evidential strength – has no plausible qualitative analog in our framework.

Our results are important for three related reasons. First, several purported counterexamples to LP are, we think, more accurately interpreted as objections to LL$^+$. We discuss one such example in §1. Because our results show that LP generalizes to qualitative settings in ways LL$^+$ may not, there is good reason to distinguish the two theses carefully.

Second, our results provide further reason to endorse Berger and Wolpert [1988, p. 141]'s conclusion that "the only satisfactory method of [statistical] analysis based on LP seems to be robust Bayesian analysis." In standard/quantitative statistical settings, only LL$^+$, we think, provides a foundation for non-Bayesian, likelihoodist techniques; the other two likelihoodist theses (i.e., LP and LL) provide little reason to use maximum-likelihood estimation or related techniques. Unfortunately, the plausibility of LL$^+$ might be a modeling artifact: when likelihood functions are assumed to be real-valued, LL$^+$ is equivalent to another plausible principle for evidential reasoning, which we call the "robust Bayesian support

(e-mail: conormw@uw.edu) (e-mail: sarafa@cs.washington.edu)

principle." That equivalence, however, breaks down in more general, qualitative settings. In such settings, the robust Bayesian favoring principle remains plausible, but LL$^+$ cannot even be articulated.

Finally, LP *sometimes* recommends against the use of p-values, confidence intervals, and other classical/frequentist summaries of data. Therefore, if LP is a special case of a more general, qualitative principle of scientific reasoning, then *some* frequentist techniques will conflict with more general principles of scientific reasoning. In the final section of the paper, we discuss one example of potentially problematic frequentist reasoning; we show that one argument for the irrelevance of stopping rules extends naturally to non-statistical data.

1. LIKELIHOODISM

At least three distinct theses are called "the likelihood principle" and are used to motivate likelihood-based methods (e.g., maximum likelihood estimation) and Bayesian tools. For clarity, we distinguish the three:

- Likelihood principle (LP): Let E and F are outcomes of two experiments $\mathbb{E}$ and $\mathbb{F}$ respectively. If there is some $c > 0$ such that $P_\theta^{\mathbb{E}}(E) = c \cdot P_\theta^{\mathbb{F}}(F)$ for all $\theta \in \Theta$, then E and F are evidentially equivalent.
- Weak Law of likelihood (LL): $P_{\theta_1}^{\mathbb{E}}(E) > P_{\theta_2}^{\mathbb{E}}(E)$ if and only if the data E favors θ_1 over θ_2.
- Strong Law of likelihood (LL$^+$): The likelihood ratio $P_{\theta_1}^{\mathbb{E}}(E)/P_{\theta_2}^{\mathbb{E}}(E)$ quantifies the degree to which the data E favors θ_1 over θ_2.

For likelihoodists (who endorse the above principles but do not embrace Bayesianism), the notions of "evidential equivalence" and "favoring" are undefined primitives that are axiomatized by principles like LP and LL. Likelihoodists, therefore, defend LP, LL, and LL$^+$ by arguing the three principles accord with (i) our informed reflections about evidential strength, (ii) accepted statistical methods that have been successfully applied in science, and (iii) other plausible and less-controversial "axioms" for statistical inference (e.g., the sufficiency principle).

Notice that LL and LL$^+$ concern only *simple/point* hypotheses (i.e., elements of Θ); the extension of those theses to *composite* hypotheses (i.e., subsets of Θ) is controversial, especially when nuisance parameters are present (see e.g., Royall [1997] [§1.7 and Chapter 7] and Bickel [2012] for different strategies). LL$^+$ is stronger than LL in two related ways: it allows one to (1) compare the *strength* of two pieces of evidence E and F by assessing whether $P_{\theta_1}^{\mathbb{E}}(E)/P_{\theta_2}^{\mathbb{E}}(E) \geq P_{\theta_1}^{\mathbb{F}}(F)/P_{\theta_2}^{\mathbb{F}}(F)$ or vice versa, and (2) compare pieces of evidence drawn from different experiments. Two examples will clarify and distinguish the theses.

Example 1: Suppose two experiments are designed to distinguish $\Theta = \{\theta_1, \theta_2\}$. The possible outcomes of the experiments (A, B, C, A', etc.) are the column headers in the tables below, and the likelihood functions are the column vectors. We intentionally list only two possible outcomes of Experiment $\mathbb{F}$.

Experiment $\mathbb{E}$

	A	B	C
θ_1	.423	.564	.011
θ_2	.039	.052	.909

Experiment $\mathbb{F}$

	A'	B'	...
θ_1	.846	.12	...
θ_2	.078	.01	...

According to LP, the outcomes A and B from $\mathbb{E}$ are evidentially equivalent to one another because $P^{\mathbb{E}}_{\theta_1}(A) = 3/4 \cdot P^{\mathbb{E}}_{\theta_1}(B)$ for all $\theta \in \{\theta_1, \theta_2\}$. Similarly, both A and B are evidentially equivalent to observing A' in $\mathbb{F}$ as $P^{\mathbb{E}}_{\theta_1}(A) = 1/2 \cdot P^{\mathbb{F}}_{\theta_1}(A')$ for all θ. However, LP says nothing about which of the hypotheses in Θ are favored by A, B, and B'. In general, LP tells one only when two samples should yield identical estimates, but it says nothing about what those estimates should be.

According to LL, the outcome A "favors" θ_1 over θ_2 because $P^{\mathbb{E}}_{\theta_1}(A) > P^{\mathbb{E}}_{\theta_2}(A)$. Similarly, LL entails that B favors θ_1 over θ_2. However, LL tells one nothing about how strong those pieces of evidence are, and in particular, whether A and B provide evidence of equal or different strength.

In contrast, LL$^+$ entails that A and B favor θ_1 to θ_2 to the same degree, and both A and B are weaker pieces of evidence than than B' because $P^{\mathbb{E}}_{\theta_1}(A)/P^{\mathbb{E}}_{\theta_2}(A) < 12 = P^{\mathbb{F}}_{\theta_1}(B')/P^{\mathbb{F}}_{\theta_2}(B')$. Notice that both LL$^+$ and LP allow one to draw conclusions about evidential value *without knowing the sample space of the second experiment.* It is for this reason that authors often stress that likelihoodist principles entail the "irrelevancy of the sample space" [Royall, 1997, §1.11].

(End Example)

Failure to distinguish the above theses carefully, we think, has caused confusion in debates about LP. For instance, [Evans et al., 1986] offer the following example as a challenge to LP; the example is due to [Fraser et al., 1984].

Example 2: Suppose $\Theta = \Omega = \mathbb{N}^+ = \{1, 2, \ldots\}$ and that P_θ assigns equal probability of 1/3 to each of the values in $\{\lfloor \theta/2 \rfloor, 2\theta, 2\theta + 1\}$. Here, $\lfloor x \rfloor$ rounds x down to the nearest integer except when $x = 1/2$, in which case $\lfloor 1/2 \rfloor = 1$. Fraser et al. [1984] note that, surprisingly, although the likelihood function is in some sense "flat", the estimator $\hat{\theta}_L(n) = \{\lfloor n/2 \rfloor\}$ gives a 2/3 confidence set, whereas the estimators $\hat{\theta}_M(n) = \{2n\}$ and $\hat{\theta}_H(\omega) = \{2n + 1\}$ each have coverage only 1/3. Evans et al. [1986] conclude, "Examples such as this would seem to make the likelihood principle questionable for statistical inference."

However, in this example, LP implies only the unobjectionable claim that each $\omega \in \Omega$ is evidentially equivalent to itself. To see why, suppose $n, m \in \Omega$ are distinct, and so we may assume $n > m$ without loss of generality. Recall that LP says that n and m are evidentially equivalent if there is some constant c such that $P_\theta(n) = c \cdot P_\theta(m)$ for *all* θ. To see that LP fails to entail that n and m are evidentially equivalent, let $\theta_0 = 2n$. Then, $P_{\theta_0}(n) = 1/3$, while $P_{\theta_0}(m) = 0$ since $m < \lfloor \theta_0/2 \rfloor = n$. Thus, there is no $c > 0$ such that $P_\theta(n) = c \cdot P_\theta(m)$ for all θ. Hence, LP says nothing about the relationship between distinct outcomes.

One might think the example challenges LL and LL$^+$. Because only $\hat{\theta}_L(n) = \{\lfloor n/2 \rfloor\}$ gives a 2/3 confidence set, one might infer that observing n favors $\lfloor n/2 \rfloor$ over $2n$ and $2n + 1$. But LL and LL$^+$ entail that n does not favor any of the parameters $\lfloor n/2 \rfloor, 2n$ or $2n + 1$ over the others. Although we think the example is not a problem for LL or LL$^+$, we will not argue so here. For now, we stress that the example shows the importance of distinguishing LP from LL$^+$ and LL.

(End Example)

The failure to distinguish LL from LP, we conjecture, is a result of assuming that the mathematical formulations of LP (like Birnbaum [1962]'s above) capture certain informal statements of the principle. Many practitioners quote Berger and Wolpert [1988, p. 1]'s informal gloss of LP; they say that LP "essentially states that all evidence, which is obtained from an experiment, about an unknown quantity θ, is contained in the likelihood function of θ for the given data." Yet other informal glosses of LP differ significantly. For example, Birnbaum [1962, p. 271] says LP asserts the "irrelevance of [experimental] outcomes not actually observed." Although these informal statements capture important insights about likelihoodist theses, we think care should be taken to stick to the technical statements when evaluating particular examples, as shown above.

Distinguishing the three theses is also important for understanding which statistical tests and techniques are justified by each. For instance, as we noted above, the formal statement of LP says nothing about which estimates are favored by a sample. LL entails that the MLE is always favored over rivals, but it does not say by how much. Thus, LL is of little use when the fit of the MLE needs to be weighed against considerations of simplicity and prior plausibility. Only LL$^+$, we think, can be used to justify likelihoodist estimation procedures, but sadly, it is also the only thesis that has no obvious qualitative analog in our framework below.

2. BAYSIANISM AND LIKELIHOODISM

LL, LP, and LL$^+$ are often said to be "compatible" with Bayes rule [Edwards, 1984, p. 28]. Here is how that "compatibility" is often explained for LL$^+$. Think of an experiment $\mathbb{E}$ as a pair $\langle \Omega^{\mathbb{E}}, \{P^{\mathbb{E}}_\theta\}_{\theta \in \Theta} \rangle$, where $\Omega^{\mathbb{E}}$ represents the possible outcomes of the experiment and $P^{\mathbb{E}}_\theta(\cdot)$ is a probability distribution over $\Omega^{\mathbb{E}}$ that specifies how likely each outcome is if θ is the true value of the parameter. Suppose Q is a prior probability distribution over Θ, which we will assume is finite for the remainder of the paper to avoid measurability assumptions. Define the posterior in the standard way:

$$Q^{\mathbb{E}}(H|E) = \frac{Q^{\mathbb{E}}(E \cap H)}{Q^{\mathbb{E}}(E)} := \frac{\sum_{\theta \in H} P^{\mathbb{E}}_\theta(E) \cdot Q(\theta)}{\sum_{\theta \in \Theta} P^{\mathbb{E}}_\theta(E) \cdot Q(\theta)}$$

for any $H \subseteq \Theta$. Then Bayes' Rule entails

$$\frac{Q^{\mathbb{E}}(\theta_1|E)}{Q^{\mathbb{E}}(\theta_2|E)} = \frac{P^{\mathbb{E}}_{\theta_1}(E)}{P^{\mathbb{E}}_{\theta_2}(E)} \cdot \frac{Q(\theta_1)}{Q(\theta_2)}$$

Although this calculation is standard in introductory remarks about Bayes rules (e.g., see [Gelman et al., 2013, p. 8]), we note that it holds *for any prior* Q. So the likelihood ratio is a measure of the degree to which *all* Bayesians' posterior degrees of belief in θ_1 increase (or decrease) upon learning E. We introduce another way of capturing that same idea below.

We would like to suggest one new argument for LL$^+$, as it will be important below. LL$^+$ *unifies* LL and LP, in the sense that it entails both theses. Why? Assume that, for every piece of evidence E and any two hypotheses θ_1 and θ_2, there is a numerical degree $\mathsf{deg}(E, \theta_1, \theta_2)$ to which E favors θ_1 over θ_2 (here, the degree might be negative). To show LL$^+$ entails LL under plausible assumptions, say that E *favors* θ_1 over θ_2 if $\mathsf{deg}(E, \theta_1, \theta_2) > \mathsf{deg}(\Omega, \theta_1, \theta_2)$, in other words, if E provides better evidence for θ_1 (over θ_2) than the sure event. If LL$^+$ holds, then $\mathsf{deg}(E, \theta_1, \theta_2) = P_{\theta_1}(E)/P_{\theta_2}(E)$, and so E favors θ_1 over θ_2 precisely if

$P_{\theta_1}(E)/P_{\theta_2}(E) = \mathsf{deg}(E, \theta_1, \theta_2) > \mathsf{deg}(\Omega, \theta_1, \theta_2) = P_{\theta_1}(\Omega)/P_{\theta_2}(\Omega) = 1$. In other words, LL^+ plus the above assumptions entails that E favors θ_1 over θ_2 precisely if $P_{\theta_1}(E) > P_{\theta_2}(E)$, exactly as LL asserts.[1]

To show LL^+ entails LP under plausible assumptions, say that E and F are *evidentially equivalent* if $\mathsf{deg}(E, \theta_1, \theta_2) = \mathsf{deg}(F, \theta_1, \theta_2)$ for all θ_1 and θ_2. In other words, E and F favor all hypotheses by equal amounts. Again, if LL^+ holds, then $\mathsf{deg}(E, \theta_1, \theta_2) = P_{\theta_1}(E)/P_{\theta_2}(E)$, and so E and F are evidentially equivalent precisely if

$$\frac{P_{\theta_1}(E)}{P_{\theta_2}(E)} = \frac{P_{\theta_1}(F)}{P_{\theta_2}(F)} \tag{1}$$

for all θ_1 and θ_2. If there is $c > 0$ such that $P_\theta(E) = c \cdot P_\theta(F)$ for all θ, then Equation 1 holds. So LL^+ entails LP.

In sum, (i) LL^+ is "compatible" with Bayes rule in the sense that the likelihood ratio is also the ratio of *all* Bayesians' posterior to prior probabilities, and (ii) LL^+ entails both LL and LP, which are also thought to be intuitively plausible (and are "compatible" with Bayesianism in still other ways). In the next section, we first prove that LL^+ is compatible with Bayes rule in another important way: it characterizes when *all* Bayesians agree that E is better evidence than F. In the second half of the paper, we show that the qualitative analog of such Bayesian agreement unifies qualitative analogs of LL and LP in the same way LL^+ unifies quantitative versions of those principles.

2.1 Robust Bayesianism and Likelihoodism

Robust Bayesianism is roughly the thesis that statistical decisions should be stable under a variety of different prior distributions [Berger, 1990, Kadane, 1984]. In this section, we discuss three elementary propositions that show likelihoodist theses characterize when all Bayesians agree about how evidence ought to change one's posterior. The propositions, we think, clarify why advocates of robust Bayesian analysis have found likelihoodist theses so attractive. The main results of our paper – in the next section – are the qualitative analogs of the second and third claim below.

Before stating our main definition, we introduce some notation. Given an experiment $\mathbb{E}$, we let $\Delta^{\mathbb{E}} = \Theta \times \Omega^{\mathbb{E}}$. We use H_1, H_2 etc. to denote subsets of Θ, and E, F, etc. to denote subsets of Ω. Again, we drop the superscript $\mathbb{E}$ when it is clear from context. We will write $Q(\cdot|H)$ and $Q(\cdot|E)$ instead of $Q(\cdot|H \times \Omega)$ and $Q(\cdot|\Theta \times E)$, and similarly for events to the left of the conditioning bar.

Definition 1. Suppose $H_1, H_2 \subseteq \Theta$ are disjoint. Let E and F be outcomes of experiments $\mathbb{E}$ and $\mathbb{F}$ respectively. Say *E Bayesian supports* H_1 over H_2 *at least as much as* F if $Q^{\mathbb{E}}(H_1|E \cap (H_1 \cup H_2)) \geq Q^{\mathbb{F}}(H_1|F \cap (H_1 \cup H_2))$ for all priors Q for which $Q^{\mathbb{F}}(\cdot|F \cap (H_1 \cup H_2))$ is well-defined. If the inequality is strict for all such Q, then we say *E Bayesian supports* H_1 over H_2 *strictly more than* F. In the former case, we write $E \,{}_{H_1}\!\overset{\mathcal{B}}{\trianglerighteq}_{H_2} F$, and in the latter, we write $E \,{}_{H_1}\!\overset{\mathcal{B}}{\gg}_{H_2} F$.

Our first goal is to find a necessary and sufficient condition for Bayesian support that involves only likelihoods. The following claim does exactly that.

[1] Notice, we drop the superscript $\mathbb{E}$ when it's clear from context.

CLAIM 1. Suppose H_1 and H_2 are finite and disjoint. Let E and F be outcomes of experiments $\mathbb{E}$ and $\mathbb{F}$ respectively. Then $E \ {}_{H_1}^{\mathcal{B}}\!\trianglerighteq_{H_2} F$ if and only if (1) for all $\theta \in H_1 \cup H_2$, if $P_\theta^{\mathbb{F}}(F) > 0$, then $P_\theta^{\mathbb{E}}(E) > 0$, and (2) for all $\theta_1 \in H_1$ and $\theta_2 \in H_2$:

$$P_{\theta_1}^{\mathbb{E}}(E) \cdot P_{\theta_2}^{\mathbb{F}}(F) \geq P_{\theta_2}^{\mathbb{E}}(E) \cdot P_{\theta_1}^{\mathbb{F}}(F)$$

Similarly, $E \ {}_{H_1}^{\mathcal{B}}\!\gg_{H_2} F$ if and only if 1 holds and the inequality in the above equation is always strict.

We omit the proof, as it uses only basic probability theory. However, details of all omitted proofs (including the elementary ones) in this paper can be found in the supplementary materials. The claim asserts that $E \ {}_{H_1}^{\mathcal{B}}\!\trianglerighteq_{H_2} F$ exactly when LL$^+$ says that E provides stronger evidence for θ_1 over θ_2 than F does. The Bayesian support relation, therefore, is one way of clarifying how a robust Bayesian can interpret the notion of "favoring" in likelihoodist theses like LL$^+$. This is important because, as we show below, the Bayesian support relation has a direct qualitative analog, unlike LL$^+$, which requires numerical degrees of favoring.

The Bayesian support relation is also important because it can be used to define the following notions of "Bayesian favoring" and "Bayesian favoring equivalence" that are respectively equivalent to the notions of "favoring" in LL and "evidential equivalence" in LP. This should be unsurprising given our argument above that LL$^+$ can be used to derive LL and LP under plausible assumptions.

DEFINITION 2. *E Bayesian favors* H_1 to H_2 if $E \ {}_{H_1}^{\mathcal{B}}\!\trianglerighteq_{H_2} \Omega$. Say it strictly does if $E \ {}_{H_1}^{\mathcal{B}}\!\gg_{H_2} \Omega$.

So Bayesian favoring is a special case of Bayesian support. In essence, the support relation *compares* (a) how much E Bayesian favors H_1 to H_2 to (b) how much F Bayesian favors H_1 to H_2. Thus, the next claim, which shows that Bayesian favoring is equivalent to LL, follows directly from Claim 1.

CLAIM 2. E Bayesian favors H_1 to H_2 if and only if (1) $P_\theta(E) > 0$ for all $\theta \in H_1 \cup H_2$ and (2) $P_{\theta_1}(E) \geq P_{\theta_2}(E)$ for all $\theta_1 \in H_1$ and $\theta_2 \in H_2$. If $H_1 = \{\theta_1\}$ and $H_2 = \{\theta_2\}$ are simple, E strictly Bayesian favors H_1 to H_2 if and only if LL entails E favors H_1 to H_2.

Next, we study two robust Bayesian notions of evidential equivalence that are easy to translate to the qualitative setting.

DEFINITION 3. Say E and F are *Bayesian posterior equivalent* if all priors Q (1) $Q^{\mathbb{E}}(\cdot|E)$ is well-defined if and only if $Q^{\mathbb{F}}(\cdot|F)$ is, and (2) $Q^{\mathbb{E}}(H|E) = Q^{\mathbb{F}}(H|F)$ for all hypotheses H and for which those conditional probabilities are well-defined.

DEFINITION 4. E and F are *Bayesian favoring equivalent* if $F \ {}_{H_1}^{\mathcal{B}}\!\trianglerighteq_{H_2} E$ and $E \ {}_{H_1}^{\mathcal{B}}\!\trianglerighteq_{H_2} F$ for all disjoint hypotheses H_1 and H_2. In other words, E and F are Bayesian favoring equivalent if for any prior Q and any disjoint hypotheses H_1 and H_2 we have (1) $Q^{\mathbb{E}}(\cdot|E \cap (H_1 \cup H_2))$ is well-defined if and only if $Q^{\mathbb{F}}(\cdot|F \cap (H_1 \cup H_2))$ is well-defined and (2) $Q^{\mathbb{E}}(H_1|E \cap (H_1 \cup H_2)) = Q^{\mathbb{F}}(H_1|E \cap (H_1 \cup H_2))$ whenever those conditional probabilities are well-defined.

It is well-known that, if LP entails E and F are evidentially equivalent, then they are posterior equivalent [Edwards et al., 1984, p. 56]; that fact makes precise in what sense LP is compatible with Bayesianism. This is one reason we conjecture that some researchers (e.g., see Example 3 of Wechsler et al. [2008]) have isolated posterior equivalence as a relation worthy of study. It is easy to show that posterior and favoring equivalence are identical relations.

Claim 3. The following are equivalent: (1) LP entails E and F are evidentially equivalent, (2) E and F are Bayesian posterior equivalent, (3) E and F are Bayesian favoring equivalent.

The three claims above show that there is an alternative way of explaining why (1) LL^+ and LP are intuitively plausible in many cases and (2) LL^+ "unifies" LL and LP. To see why, consider one last philosophical thesis about statistical evidence.

Robust Bayesian Support Principle: E provides at least as good statistical evidence for H_1 over H_2 as F if $E \,{}_{H_1}\!\overset{\mathcal{B}}{\trianglerighteq}_{H_2} F$.

By Claim 1, the robust Bayesian support principle is equivalent to LL^+. And by Claim 2 and Claim 3, it entails LL and LP with some plausible, additional assumptions, thereby "unifying" LL and LP in roughly the same way LL^+ does. Thus, LL^+ may be plausible only because it is equivalent to the robust Bayesian support principle in common statistical settings. And the robust Bayesian support principle is plausible, we conjecture, precisely because it captures the idea that evidence persuades *all* rational parties to change their beliefs in particular ways. This is important because, as we now show, the robust Bayesian principle generalizes to qualitative settings in which LL^+ cannot be formulated.

3. QUALITATIVE LIKELIHOODISM

3.1 Key Concepts

To move from quantitative to qualitative probability, we replace probability functions with two *orderings*. As before, let Θ be the set of simple hypotheses, and for any experiment $\mathbb{E}$, we let $\Omega^{\mathbb{E}}$ the set of experimental outcomes. The first relation, $\sqsubseteq^{\mathbb{E}}$, is the qualitative analog of the set of likelihood functions. As before, we drop $\mathbb{E}$ when it's clear from context. Informally, $A|\theta \sqsubseteq B|\eta$ represents the claim that "experimental outcome B is at least as likely under supposition η as outcome A is under supposition θ"; it is the qualitative analog of $P_\theta(A) \leq P_\eta(B)$. We write $A|\theta \equiv B|\theta$ if $A|\theta \sqsubseteq B|\theta$ and vice versa.

Although we typically consider expressions of the form $A|\theta \sqsubseteq B|\eta$, the $\sqsubseteq$ relation is also defined when experimental outcomes $E \subseteq \Omega$ appear to the right

Quantitative Probabilistic Notions	**Qualitative Analog**
$P^{\mathbb{E}}_\theta(E) \leq P^{\mathbb{F}}_\upsilon(F)$	$E\|\theta \sqsubseteq F\|\upsilon$
$Q^{\mathbb{E}}(H_1\|E) \leq Q^{\mathbb{F}}(H_2\|F)$	$H_1\|E \preceq H_2\|F$
LL $\Leftrightarrow$ E Bayesian favors H_1 to H_2	QLL $\Leftrightarrow$ E qualitatively favors H_1 to H_2.
LP $\Leftrightarrow$ E and F are Bayesian posterior and favoring equivalent	QLP^- $\Rightarrow$ E and F are qualitatively posterior and favoring equivalent.

of the conditioning bar, i.e., $A|\theta \cap E \sqsubseteq A|\theta \cap F$ is a well-defined expression if $E, F \subseteq \Omega$. However, it is not-well defined when composite hypotheses H appear to the right of the conditioning bar, just as there are no likelihood functions $P_H(\cdot)$ for composite hypotheses in the quantitative case.

Bayesians assume that beliefs are representable by a probability function. We weaken that assumption and assume beliefs are representable by an ordering $\preceq$ on $\mathcal{P}(\Delta) \times \mathcal{P}(\Delta)$, where $\mathcal{P}(\Delta)$ is the power set of Δ. As before, if $E \subseteq \Omega$ and $H \subseteq \Theta$ we write $E|H$ instead of $\Theta \times E|H \times \Omega$. In the special case in which $H = \{\theta\}$ is a singleton, we omit the curly brackets and write $E|\theta$ instead of $E|\{\theta\}$. We write $A|B \sim C|D$ is $A|B \preceq C|D$ and vice versa.

The notation is suggestive; $A|B \preceq C|D$ if the experimenter regards C as at least as likely under supposition D as A would be under supposition B. Just as Bayesians condition on both composite hypotheses and experimental outcomes, there are no restrictions on what can appear to the right of the conditioning bar in expressions involving $\preceq$.

Clearly, to prove anything, we must assume that $\sqsubseteq$ and $\preceq$ satisfy certain constraints; those constraints are specified in Section 3.3. But before stating those constraints, we state our main results.

3.2 Main Results

It is now easy to state a qualitative analog of LL:

- Qualitative law of likelihood (QLL): E favors θ_1 to θ_2 if $E|\theta_2 \sqsubset E|\theta_1$.

Claim 2 asserts that LL characterizes precisely when E Bayesian favors θ_1 over θ_2, i.e., when all Bayesian agents agree E favors one hypothesis over another. So by analogy to the definition of. “Bayesian support” and “Bayesian favoring”, define:

Definition 5. *E qualitatively favors* H_1 to H_2 at least as much as F if $H_1|E \cap (H_1 \cup H_2) \succeq H_1|F \cap (H_1 \cup H_2)$ for all orderings $\preceq$ satisfying the axioms below and for which the expression $\cdot|F \cap (H_1 \cup H_2)$ is well-defined.

Definition 6. *E qualitatively favors* H_1 to H_2 if E supports H_1 to H_2 at least as much as Ω.

Our first major result is the qualitative analog of Claim 2. Just as Claim 2 entails that E Bayesian favors θ_1 to θ_2 when LL entails so, our first major result shows that E qualitatively favors θ_1 to θ_2 if and only if QLL entails so. Under mild assumptions, this equivalence can be extended to finite composite hypotheses.

Theorem 1. Suppose H_1 and H_2 are finite. Then E qualitatively favors H_1 over H_2 if (1) $\emptyset|\theta \sqsubset E|\theta$ for all $\theta \in H_1 \cup H_2$ and (2) $E|\theta_2 \sqsubseteq E|\theta_1$ for all $\theta_1 \in H_1$ and $\theta_2 \in H_2$. Under Assumption 1 (below), the converse holds as well. In both directions, the favoring inequality is strict exactly when the likelihood inequality is strict. If $H_1 = \{\theta_1\}$ and $H_2 = \{\theta_2\}$ are simple, then E qualitatively favors H_1 over H_2 if and only if QLL entails so. No additional assumptions are required in this case.

Assumption 1 below more-or-less says that one's "prior" ordering over the hypotheses is unconstrained by the "likelihood" relation $\sqsubseteq$. This is exactly analogous to the quantitative probability theory. In the quantitative case, a joint distribution $Q^{\mathbb{E}}$ on $\Delta = \Theta \times \Omega^{\mathbb{E}}$ is determined (i) the measures $\langle P_\theta(\cdot)\rangle_{\theta\in\Theta}$ over experimental outcomes $\Omega^{\mathbb{E}}$ and (ii) one's prior Q over the hypotheses Θ. Although the joint distribution $Q^{\mathbb{E}}$ is constrained by $\langle P_\theta(\cdot)\rangle_{\theta\in\Theta}$, one's prior Q on Θ is not, and so for any non-empty hypothesis $H \subseteq \Theta$, one can define some Q such that $Q(H) = 1$ and $Q(\theta) > 0$ for all $\theta \in H$. That's what the following assumption says in the qualitative case.

Assumption 1. For all orderings $\sqsubseteq$ satisfying the axioms above and for all non-empty $H \subseteq \Theta$, there exists an ordering $\preceq$ satisfying the axioms such that (A) $H|\Delta \sim \Delta|\Delta$ and (B) $\theta|\Delta \succ \emptyset|\Delta$ for all $\theta \in H$.

We believe Assumption 1 is provable, but we have not yet produced a proof.

Our second main result is a qualitative analog of Claim 3, which says that LP characterizes when two pieces of evidence are posterior and Bayesian-favoring equivalent. The qualitative analogs of posterior equivalence and favoring equivalence are obvious, but we state them for full clarity.

Definition 7. E and F are *qualitative favoring equivalent* if E supports H_1 over H_2 at least as much as F and vice versa, for any two disjoint hypotheses H_1 and H_2.

Definition 8. E and F are *qualitative posterior equivalent* if all orderings $\preceq$ that satisfy the axioms (1) the expression $\cdot|E$ is well-defined if and only if $\cdot|E$ is, and (2) $H|E \sim H|F$ for all hypotheses $H \subseteq \Theta$.

As before, these two definitions are equivalent.

Claim 4. *E and F are qualitative posterior equivalent if and only if they are qualitative favoring equivalent.*

Although posterior and favoring equivalence are easy to make qualitative, what about LP? Quantitatively, the statement of LP mentions both multiplication and a global constant $c > 0$, which seem difficult to translate to the qualitative setting.

Notice, however, that we can multiply probabilities when two events are independent. So suppose E and F are outcomes of the same experiment and

1. $P_\theta(E) = P_\theta(F \cap C_\theta)$ for all θ,
2. For all $\theta \in \Theta$, the events F and C_θ are conditionally independent given θ, and
3. $P_\theta(C_\theta) = P_\upsilon(C_\upsilon) > 0$ for all $\theta, \upsilon \in \Theta$.

Roughly, the event C_θ acts as a witness to the equality $P_\theta(E) = c \cdot P_\theta(F)$. Specifically, assumptions 1 and 2 encode the equality, and assumption 3 asserts this constant is invariant with respect to the parameter θ.

	Number of balls				
	Blue	White	Cyan	Cobalt	Total
Urn Type 1	15	30	50	5	100
Urn Type 2	10	20	20	50	100

TABLE 1

Drawing a blue ball is evidentially equivalent to drawing a white ball according to LP. *In this example, the equivalence can be "encoded" by the cyan and cobalt balls instead of a constant, c.*

By LP, the three conditions entail that E and F are evidentially equivalent. The proof is simple. Let $c = P_{\theta_0}(C_{\theta_0})$ for any $\theta_0 \in \Theta$. Then for all θ:

$$\begin{aligned} P_\theta(E) &= P_\theta(F \cap C_\theta) \text{ by Assumption 1} \\ &= P_\theta(F) \cdot P_\theta(C_\theta) \text{ by Assumption 2} \\ &= c \cdot P_\theta(F) \text{ by Assumption 3} \end{aligned}$$

Since $P_\theta(E) = c \cdot P_\theta(F)$ for all θ, then E and F are evidentially equivalent by LP.

For an example, suppose we are trying to discern the type of an unmarked urn, with colored balls in frequencies according to Table 1. Let B and W respectively denote the events that one draws a blue and a white ball on the first draw. By LP, these events are equivalent, as $P_\theta(B) = 1/2 \cdot P_\theta(W)$ for all $\theta \in \{\theta_1, \theta_2\}$, which denote the urn type. We can see this equivalence in another way. Let C_1/C_2 respectively denote the events that, after replacing the first draw, one draws a cyan/cobalt ball respectively on the second draw. Then, $P_{\theta_1}(W \cap C_1) = P_{\theta_1}(W) \cdot P_{\theta_1}(C_1) = 1/2 \cdot P_{\theta_1}(W) = P_{\theta_1}(B)$ by independence of the draws, if the urn is Type 1. Similarly, $P_{\theta_2}(W \cap C_2) = P_{\theta_2}(W) \cdot P_{\theta_2}(C_2) = 1/2 \cdot P_{\theta_2}(W) = P_{\theta_2}(B)$, if the urn is Type 2. By LP, because $P_{\theta_1}(C_1) = P_{\theta_2}(C_2) > 0$ and $P_{\theta_i}(B) = P_{\theta_i}(W \cap C_i) = P_{\theta_i}(C_i) \cdot P_{\theta_i}(W)$ for all i, we know B and W are evidentially equivalent.

So define LP$^-$ to be the thesis that, if conditions 1-3 hold, then E and F are evidentially equivalent. We just showed that, if LP$^-$ entails E and F are evidentially equivalent, then so does LP. By Claim 3, it follows that E and F are also Bayesian posterior and favoring equivalent. Our second major result is the qualitative analog of that fact.

Because conditions 1-3 do not contain any arithmetic operations, each has a direct qualitative analog. The only slightly tricky condition is the second. To define a qualitative analog of conditional independence, note that two events A and B are conditionally independent given C if and only if $P(A|B \cap C) = P(A|C)$ or $P(B \cap C) = 0$. Analogously, we will say events A and B are qualitatively conditionally independent given C if $A|B \cap C \sim A|C$ or $B \cap C \in \mathcal{N}$. In this case, we write $A \perp\!\!\!\perp_C B$.[2]

Our second main result is the following:

THEOREM 2. Let $\{C_\theta\}_{\theta \in \Theta}$ be events such that

1. $E|\theta \equiv F \cap C_\theta|\theta$ for all $\theta \in \Theta$,
2. $F \perp\!\!\!\perp_\theta C_\theta$ for all $\theta \in \Theta$, and
3. $\emptyset|\theta \sqsubset C_\theta|\theta \equiv C_\eta|\eta$ for all $\theta, \eta \in \Theta$.

[2]The lemmas below show this definition of (qualitative) independence has the desired properties. For example, Lemma 3 entails $A \perp\!\!\!\perp_C B$ if and only if $B \perp\!\!\!\perp_C A$.

If Θ is finite, then E and F are qualitatively posterior and favoring equivalent.

The following corollary of Theorem 2 is the qualitative analog of the fact that when $P_\theta(E) = P_\theta(F)$ (i.e., when the constant c in LP equals one), then E and F are posterior and favoring equivalent. In the quantitative case, the corollary is trivial because one can let $C_\theta = \Delta$ for all θ. The three conditions of Theorem 2 are satisfied then since (1) $P_\theta(E) = P_\theta(F) = P_\theta(F \cap \Delta)$ for all θ by assumption, (2) F and Δ are conditionally independent given every θ, as $P_\theta(F \cap \Delta) = P_\theta(F) = P_\theta(F) \cdot 1 = P_\theta(F) \cdot P_\theta(\Delta)$, and (3) $P_\theta(\Delta) = 1 > 0$ for all $\theta \in \Theta$. Analogous reasoning works in the qualitative case using the axioms below.

COROLLARY 1. If Θ is finite and $E|\theta \equiv F|\theta$, then E and F are qualitatively posterior and favoring equivalent.

3.3 Axioms for Qualitative Probability

We assume that both $\sqsubseteq$ and $\preceq$ satisfy the first set of axioms below. To state the axioms, therefore, we let $\underline{\blacktriangleleft}$ be either $\sqsubseteq$ or $\preceq$, and let $\blacktriangleleft$ be the corresponding "strict" inequality defined by $x \blacktriangleleft y$ if $x \underline{\blacktriangleleft} y$ and $y \not\underline{\blacktriangleleft} x$. Define $A|B \overset{\blacktriangle}{=} C|D$ if and only if $A|B \underline{\blacktriangleleft} C|D$ and vice versa.

> Axiom 1: $\underline{\blacktriangleleft}$ is a weak order (i.e., it is linear/total, reflexive, and transitive).
> Axiom 3: $A|A \overset{\blacktriangle}{=} B|B$ and $A|B \underline{\blacktriangleleft} \Delta|C$ for all $C \notin \mathcal{N}$.
> Axiom 4: $A \cap B|B \overset{\blacktriangle}{=} A|B$.
> Axiom 5: Suppose $A \cap B = A' \cap B' = \emptyset$. If $A|C \underline{\blacktriangleleft} A'|C'$ and $B|C \underline{\blacktriangleleft} B'|C'$, then $A \cup B|C \underline{\blacktriangleleft} A' \cup B'|C'$; moreover, if either hypothesis is $\blacktriangleleft$, then the conclusion is $\blacktriangleleft$.
> Axiom 6: Suppose $C \subseteq B \subseteq A$ and $C' \subseteq B' \subseteq A'$.
>
> > Axiom 6a: If $B|A \underline{\blacktriangleleft} C'|B'$ and $C|B \underline{\blacktriangleleft} B'|A'$, then $C|A \underline{\blacktriangleleft} C'|A'$; moreover, if either hypothesis is $\blacktriangleleft$, the conclusion is $\blacktriangleleft$.
> >
> > Axiom 6b: If $B|A \underline{\blacktriangleleft} B'|A'$ and $C|B \underline{\blacktriangleleft} C'|B'$, then $C|A \underline{\blacktriangleleft} C'|A'$; moreover, if either hypothesis is $\blacktriangleleft$ and $C \notin \mathcal{N}$, the conclusion is $\blacktriangleleft$.

We discuss Axiom 2 below, as it is different for $\sqsubseteq$ and $\preceq$. The above axioms are due to [Krantz et al., 2006, p. 222] and enumerated in the same order as in that text. What we call Axiom 6b is what they call Axiom 6' (p. 227). For several reasons (e.g., there is no Archimedean condition), our axioms are not sufficient for $\preceq$ to be representable as either a (conditional) probability measure or a set of probability measures. See Alon and Lehrer [2014] for a recent representation theorem for sets of probability measures.

Axioms 1, 3, 4, and 5, are fairly analogous to facts of quantitative probability. Axiom 6 is useful because it allows us to "multiply" in a qualitative setting. To see its motivation, consider Axiom 6a and note that if $C \subseteq B \subseteq A$ and $C' \subseteq B' \subseteq A'$, then

$$P(C|B) = \frac{P(C)}{P(B)} \text{ and } P(B|A) = \frac{P(B)}{P(A)},$$

and similarly for the A', B', and C'. So if $P(B|A) \geq P(C'|B')$ and $P(C|B) \geq P(B'|A')$, then

$$\frac{P(B)}{P(A)} \geq \frac{P(C')}{P(B')} \text{ and } \frac{P(C)}{P(B)} \geq \frac{P(B')}{P(A')}.$$

When we multiply the left and right-hand sides of those inequalties, we obtain $P(C)/P(A) \geq P(C')/P(A')$, which is equivalent to $P(C|A) \geq P(C'|A')$ given our assumption about the nesting of the sets. Axiom 6b can be motivated similarly.

In addition to these axioms, in statistical contexts, one typically assumes that experimenters agree upon the likelihood functions of the data, which means that, in discrete contexts, $Q^{\mathbb{E}}(\cdot|\theta) = P^{\mathbb{E}}_{\theta}(\cdot)$ whenever $Q(\theta) > 0$. Similarly, we assume that $\preceq$ extends $\sqsubseteq$ in the following sense:

> Axiom 0: If $B, D \notin \mathcal{N}_{\preceq}$ and $B, D \in \Theta \times \mathcal{P}(\Omega)$, then $A|B \preceq C|D$ if and only if $A|B \sqsubseteq C|D$.

We now return to Axiom 2, which concerns probability zero events. For any given fixed θ, the conditional probability $P_\theta(\cdot|E)$ is undefined if and only if $P_\theta(E) = 0$. Similarly, we define $\mathcal{N}_{\sqsubseteq}$ to be all and only sets of that form $\{\theta\} \times E$ such that $E|\theta \sqsubseteq \emptyset|\theta$. We call such events $\sqsubseteq$-*null*. Notice that an experimenter may have a prior that assigns a parameter $\theta \in \Theta$ zero probability, even if $P_\theta(E) > 0$ for all events E. So the set of null events, in the experimenter's joint distribution, will contain all of the null events for each P_θ plus many others. Accordingly, we assume the following about $\preceq$-null events.

> Axiom 2 (for $\sqsubseteq$): $\theta \times \Omega \notin \mathcal{N}_{\sqsubseteq}$ for all θ, and $A \in \mathcal{N}_{\sqsubseteq}$ if and only if $A = \theta \times E$ and $E|\theta \sqsubseteq \emptyset|\theta$.
> Axiom 2 (for $\preceq$): $\Delta \notin \mathcal{N}$, and $A \in \mathcal{N}$ if and only if $A|\Delta \preceq \emptyset|\Delta$.

We say an expression of the form $\cdot|A$ is *undefined* with respect to $\preceq$ / $\sqsubseteq$ if A is null with respect to the appropriate relation. Because it is typically clear from context, we do not specify with respect to which ordering an expression is undefined.

3.4 Qualitative Law of Likelihood

Over the next two sections, we sketch proofs of Theorem 1 and Theorem 2 so the reader can see how our axioms work in ways analogous to quantitative probability theory. For ease of reading, we have stated all the propositions and lemmata below without qualification, even though several require the assumption that one or more events are non-null. Again, for detailed proofs that handle the case of null events, see the supplementary materials.

Our goal in this section is to sketch a proof of Theorem 1 specifically. The theorem follows immediately from the following two propositions:

PROPOSITION 1. Suppose $H_1 \cap H_2 = \emptyset$. Then $E|H_1 \succeq E|H_2$ if and only if $H_1|E \cap (H_1 \cup H_2) \succeq H_1|H_1 \cup H_2$. Further, if either side of the biconditional contains a strict inequality $\succ$, then so does the other.

PROPOSITION 2. Suppose H_1 and H_2 are finite and that $H_1 \cap H_2 = \emptyset$. Then $E|H_2 \preceq E|H_1$ for all orderings $\preceq$ (satisfying the axioms above) if $E|\theta_2 \sqsubseteq E|\theta_1$ for all $\theta_1 \in H_1$ and $\theta_2 \in H_2$. Further, under Assumption 1, if $E|H_2 \preceq E|H_1$ for all orderings $\preceq$, then $E|\theta_2 \sqsubseteq E|\theta_1$ for all $\theta_1 \in H_1$ and $\theta_2 \in H_2$.

In both directions, the left-hand side contains the strict inequality $\prec$ if and only if the right-hand side contains the strict relation $\sqsubset$.

Note that if H_1 and H_2 are simple hypotheses, i.e. $H_1 = \theta_1, H_2 = \theta_2$, then we get that $E|H_1 \preceq E|H_2$ if and only if $E|\theta_1 \sqsubseteq E|\theta_2$ by Axiom 0. The additional assumption is required when H_1 or H_2 are composite.

The proofs of these propositions require the following lemmata; the first two are analogs of the fact that $P(E|H_1 \cup H_2)$ must be between $P(E|H_1)$ and $P(E|H_2)$.

Lemma 1. Suppose $H_1 \cap H_2 = \emptyset$. Then, $E|H_1 \succeq E|H_2$ if and only if $E|H_1 \succeq E|(H_1 \cup H_2)$. Further, if either side of the biconditional is strict, the other side is strict too.

Lemma 2. Suppose $H_1 \cap H_2 = \emptyset$.

1. If $E|H_1, E|H_2 \preceq E|H_3$, then $E|(H_1 \cup H_2) \preceq E|H_3$.
2. If $E|H_3 \preceq E|H_1, E|H_2$, then $E|H_3 \preceq E|(H_1 \cup H_2)$.

If the premise is not strict *and* $E|H_1 \sim E|H_2$, then the conclusion is not strict. Otherwise, the conclusion is strict.

Finally, we often use the following variant of Axiom 6a:

Lemma 3. Suppose $A \supseteq B \supseteq C$ and $A \supseteq B' \supseteq C$. If $B|A \succeq C|B'$ and $B \notin \mathcal{N}_{\preceq}$, then $B'|A \succeq C|B$. Further, if the antecedent is $\succ$, the consequent is $\succ$.

Proof of Proposition 1: First we prove the left-to-right direction. Suppose $E|H_1 \succeq E|H_2$, where $H_1 \cap H_2 = \emptyset$. By Lemma 1, we get $E|H_1 \succeq E|(H_1 \cup H_2)$. Applying Axiom 4 to both sides of that inequality yields

$$E \cap H_1|H_1 \succeq E \cap (H_1 \cup H_2)|H_1 \cup H_2 \tag{2}$$

Define:

$$\begin{aligned} &A = H_1 \cup H_2 \\ &B = E \cap (H_1 \cup H_2) \qquad B' = H_1 \\ &C = E \cap H_1 \end{aligned}$$

Note that equation (2) says that $C|B' \succeq B|A$. So Lemma 3 tells us that $C|B \succeq B'|A$, i.e., $E \cap H_1|E \cap (H_1 \cup H_2) \succeq H_1|H_1 \cup H_2$ Applying Axiom 4 to the left-hand side of that inequality yields the desired result. Note that if we had assumed $E|H_1 \succ E|H_2$, then our conclusion would contain $\succ$ because both Lemma 3 and Lemma 1 yield strict comparisons.

In the right to left direction, suppose $H_1|E \cap (H_1 \cup H_2) \succeq H_1|H_1 \cup H_2$. By Axiom 4:

$$\begin{aligned} &H_1|E \cap (H_1 \cup H_2) \\ &\quad \sim H_1 \cap E \cap (H_1 \cup H_2)|E \cap (H_1 \cup H_2) \\ &\quad \sim E \cap H_1|E \cap (H_1 \cup H_2) \end{aligned}$$

So, we know $E \cap H_1|E \cap (H_1 \cup H_2) \succeq H_1|H_1 \cup H_2$. Now we apply Lemma 3 to

$$\begin{aligned} &A = H_1 \cup H_2 \\ &B = H_1 \qquad\qquad B' = E \cap (H_1 \cup H_2) \\ &C = E \cap H_1 \end{aligned}$$

The containment relations are satisfied, and $C|B' \succeq B|A$. Thus, we get $C|B \succeq B'|A$, i.e., $E \cap H_1|H_1 \succeq E \cap (H_1 \cup H_2)|(H_1 \cup H_2)$. Applying Axiom 4 to both sides of the last inequality yields $E|H_1 \succeq E|H_1 \cup H_2$. So by Lemma 1, we get $E|H_1 \succeq E|H_2$, as desired. As before, if the premise were $\succ$, the conclusion would also be $\succ$ because the necessary lemmas would yield strict comparisons. $\square$

Proof of Proposition 2: In the right-to-left direction, suppose $E|\theta_2 \sqsubseteq E|\theta_1$ for all $\theta_1 \in H_1$ and $\theta_2 \in H_2$. We want to show $E|H_2 \preceq E|H_1$ for all orderings $\preceq$ satisfying the axioms. So let $\preceq$ be any such ordering. We show $E|H_2 \preceq E|H_1$ by induction on the maximum of the number of elements in H_1 or H_2.

In the base case, suppose $H_1 = \{\theta_1\}$ and $H_2 = \{\theta_2\}$ both have one element. Then by Axiom 0, it immediately follows that $E|H_2 \preceq E|H_1$.

For the inductive step, suppose the result holds for all natural numbers $m \leq n$, and assume that $H_1 = \{\theta_{1,1}, \dots \theta_{1,k}\}$ and $H_2 = \{\theta_{2,1} \dots, \theta_{2,l}\}$ where either k or l (or both) is equal to $n+1$. Define $H_1' = \{\theta_{1,1}\}$ and $H_1'' = H_1 \setminus H_1'$, and similarly, define $H_2' = \{\theta_{2,1}\}$ and $H_2'' = H_2 \setminus H_2'$. Then H_1', H_1', H_2', and H_2'' all have n or fewer elements, and by assumption, $E|\theta_2 \sqsubseteq E|\theta_1$ for all $\theta_2 \in H_2', H_2''$ and all $\theta_1 \in H_1', H_1''$. By inductive hypothesis, it follows that $E|H_1', E|H_1'' \preceq E|H_2', E|H_2''$. By repeated application of Lemma 2, it follows that $E|H_1 = E|(H_1' \cup H_1'') \preceq E|(H_2' \cup H_2'') = E|H_2$.

Note that if the premise were $\sqsubset$, Axiom 0 would prove the strict version of the base case. In the inductive step, every inequality would also be strict because Lemma 2 preserves the strictness. Thus, we would get $E|H_2 \prec E|H_1$.

In the left-to-right direction, suppose $E|H_2 \preceq E|H_1$ for all orderings $\preceq$. We want to show that $E|\theta_2 \sqsubseteq E|\theta_1$ for all $\theta_1 \in H_1$ and $\theta_2 \in H_2$. So fix $\theta_1 \in H_1$ and $\theta_2 \in H_2$. We must show $E|\theta_2 \sqsubseteq E|\theta_1$. Suppose for the sake of contradiction that $E|\theta_2 \not\sqsubseteq E|\theta_1$, and thus, by totality of $\sqsubseteq$, we know $E|\theta_1 \sqsubset E|\theta_2$. We must find at least one ordering $\preceq$ that (i) satisfies Axioms 0-6 and (ii) entails $E|H_2 \not\preceq E|H_1$.

Using Assumption 1 with $H = \{\theta_1, \theta_2\}$, define $\preceq$ so that it satisfies all of the axioms and:

$$\begin{array}{rcl} \theta_1|\Delta, \theta_2|\Delta & \succ & \emptyset|\Delta \\ A|\Delta & \sim & \emptyset|\Delta \text{ if } \theta_1, \theta_2 \notin A \end{array}$$

Because $\theta_1|\Delta, \theta_2|\Delta \succ \emptyset|\Delta$, Axiom 2 entails that $\{\theta_1\}, \{\theta_2\} \notin \mathcal{N}$. Since $\{\theta_1\}, \{\theta_2\} \notin \mathcal{N}$ and $E|\theta_1 \sqsubset E|\theta_2$, by Axiom 0 we obtain that $E|\theta_1 \prec E|\theta_2$.

To finish the proof, we need one final lemma, which is the analog of the claim that $P(A|B \cup C) = P(A|B)$ whenever $P(B) > P(C) = 0$.

Lemma 4. Suppose $C \in \mathcal{N}$ and $B \notin \mathcal{N}$. Then $A|B \sim A|B \cup C$ for all A.

Now consider $H_1' = H_1 \setminus \{\theta_1\}$ and $H_2' = H_2 \setminus \{\theta_2\}$. Since H_1 and H_2 are disjoint, neither H_1' nor H_2' contain either θ_1 or θ_2. Thus by construction of $\preceq$, it follows that $H_1', H_2' \in \mathcal{N}$. So by Lemma 4, we obtain that

$$\begin{array}{rcl} E|H_1 & = & E|H_1' \cup \{\theta_1\} \sim E|\theta_1 \text{ and} \\ E|H_2 & = & E|H_2' \cup \{\theta_2\} \sim E|\theta_2 \end{array}$$

Since $E|\theta_1 \prec E|\theta_2$, it follows that $E|H_1 \prec E|H_2$, as desired.

If the premise were $\prec$, the conclusion would clearly be $\sqsubset$, by simply swapping all strict and weak inequalities in the preceding argument.

$\square$

3.5 A Qualitative Likelihood Principle

In this section, we prove Theorem 2. Let $A_\theta = F \cap C_\theta \cap \theta | \bigcup_{\eta \in \Theta} F \cap C_\eta \cap \eta$. We show that $\theta|E \sim A_\theta$ and $\theta|F \sim A_\theta$ for all θ. Hence, $\theta|E \sim \theta|F$ for all θ. Using finite additivity, we conclude $H|E \sim H|F$ for all finite hypotheses H. So E and F are qualitatively posterior equivalent if Θ is finite.

We start with, roughly speaking, a lemma that allows us to combine applications of Bayes' theorem and then the Law of Total Probability in the qualitative setting. Suppose that $B_1 \ldots B_n$ partition G and $P(A|B_i) = P(C|B_i)$ for all i. By Bayes' theorem and the Law of Total Probability, we get that $P(B_i|A \cap G) = P(B_i|C \cap G)$ for all i. The following is a qualitative analog, except we let C vary with i.

Lemma 5. Suppose $B_1, \ldots, B_n$ partition G. Further, suppose $A|B_i \sim C_i|B_i$ for all $i \leq n$. Then, $B_i|A \cap G \sim B_i \cap C_i | \bigcup_{j \leq n} B_j \cap C_j$ for all $i \leq n$.

The proof is fairly involved but unenlightening, so we omit it here.

Proposition 3. Suppose $E|\theta \equiv F \cap C_\theta|\theta$ for all $\theta \in \Theta$. If Θ is finite, then $\theta|E \sim F \cap C_\theta \cap \theta | \bigcup_{\eta \in \Theta} F \cap C_\eta \cap \eta$ for all $\theta \in \Theta$.

Proof. We apply Lemma 5. Let $\Theta = \{\theta_1, \ldots, \theta_n\}$ and $B_i = \theta_i$. Thus, $G = \Theta$, which, recall, is shorthand for $\Theta \times \Omega = \Delta$ in this context. Let $A = E$ and $C_i = F \cap C_{\theta_i}$. Then, $A|B_i \sim C_i|B_i$ is the premise of this lemma, and $B_i|A \sim B_i \cap C_i | \bigcup_{j \leq n} B_j \cap C_j$ is the conclusion. $\square$

The following lemma also involves a qualitative application of the Law of Total Probability, this time alongside the definition of conditional independence. Suppose that $B_1, \ldots, B_n$ partition G, that $A \perp\!\!\!\perp_{B_i} C_i$, and that $P(C_i|B_i) = c$ for all i. By applying the Law of Total Probability, we get that $P(\cup_i C_i \cap B_i | A \cap G) = \sum_{j \leq n} P(C_j|A \cap B_j) \cdot P(B_j|A \cap G)$. By using conditional independence and the fact that $P(C_j|B_j)$ is a constant c, we get $\sum_{j \leq n} P(C_j|A \cap B_j) \cdot P(B_j|A \cap G) = c$. The following is the qualitative analog of this argument.

Lemma 6. Suppose $B_1, \ldots, B_n$ partition G. Further, suppose $A \perp\!\!\!\perp_{B_i} C_i$ and for all $i, j \leq n : C_i|B_i \sim C_j|B_j$. Then $\bigcup_{i \leq n} C_i \cap B_i | A \cap G \sim C_k|B_k$ for all $k \leq n$.

The proof is by induction on the size of G.

Proposition 4. Suppose $F \perp\!\!\!\perp_\theta C_\theta$ and $C_\theta|\theta \equiv C_\eta|\eta$ for all $\theta, \eta \in \Theta$. If Θ is finite, then $\theta|F \sim F \cap C_\theta \cap \theta | \bigcup_{\eta \in \Theta} F \cap C_\eta \cap \eta$ for all $\theta \in \Theta$.

Proof. Let $\Theta = \{\theta_1, \ldots, \theta_n\}$ and let $B_i = \{\theta_i\}$. Thus, $G = \Theta$. Let $A = F$ and $C_i = C_\theta$. Then, from Lemma 6 we get $\bigcup_{\eta \in \Theta} C_\eta \cap \eta | F \sim C_\theta|\theta$ for all $\theta \in \Theta$. Now, since $F \perp\!\!\!\perp_\theta C_\theta$, we can change this to:

$$\bigcup_{\eta \in \Theta} C_\eta \cap \eta | F \sim C_\theta | F \cap \theta \tag{3}$$

Now, we apply Lemma 3 with:

$$
\begin{aligned}
&A = F \\
&B = \bigcup_{\eta \in \Theta} F \cap C_\eta \cap \eta \qquad B' = F \cap \theta \\
&C = F \cap C_\theta \cap \theta
\end{aligned}
$$

Equation 3 says $B|A \sim C|B'$. So by Lemma 3, we get $B'|A \sim C|B$, which is the desired conclusion. □

We have shown that the three conditions of Theorem 2 entail that E and F are qualitatively posterior equivalent if Θ is finite; by Claim 4, we get that they are qualitative favoring equivalent as well.

4. COMPARING EXPERIMENTS: STOPPING RULES AND MIXTURES

One of the most controversial consequences of LP is that it entails the "irrelevance of stopping rules" (e.g., see Savage's contribution to Savage et al. [1962]). However, the qualitative version of LP we have stated, one might argue, has no such implication. Why? The careful reader will have noticed that the relations $\sqsubseteq$ and $\preceq$ are always implicitly indexed to a fixed experiment. For example, we defined $\preceq$ to be an ordering on $\mathcal{P}(\Delta) \times \mathcal{P}(\Delta)$, where $\Delta = \Theta \times \Omega^{\mathbb{E}}$ for some fixed experiment $\mathbb{E}$. Thus, on first glance, our qualitative "analog" of LP does not allow one to compare the likelihoods of results drawn from different experiments, and such a comparison is precisely what is at issue in debates about stopping rules.

In this section, we argue that our results do not have this limitation, at least for an important class of "non-informative" stopping rules [Raiffa and Schlaifer, 1961]. To do so, let us first clarify what it means for a stopping rule to be "irrelevant." From the robust Bayesian perspective, the choice between two stopping rules $s_{\mathbb{E}}$ and $s_{\mathbb{F}}$ is typically called "irrelevant" if, whenever an outcome ω can be obtained from two experiments $\mathbb{E}$ and $\mathbb{F}$ that *differ only* in their respective stopping rules $s_{\mathbb{E}}$ and $s_{\mathbb{F}}$ respectively, it follows that $Q^{\mathbb{E}}(H|\omega) = Q^{\mathbb{F}}(H|\omega)$ for any hypothesis $H \subseteq \Theta$ and any prior Q. In other words, the stopping rule is irrelevant if, for any outcome ω that can be obtained in both experiments, observing ω in $\mathbb{E}$ is posterior equivalent to observing ω in $\mathbb{F}$. Below, we say what it means for two experiments to "differ only" in their stopping rules.

To generalize this idea to the qualitative setting, we characterize irrelevance in a second, equivalent way. Imagine a *mixed experiment* is conducted in which a fair coin is flipped to decide whether to conduct experiment $\mathbb{E}$ or $\mathbb{F}$. Let $\langle \mathbb{E}, \omega \rangle$ denote the outcome of $\mathbb{M}$ in which the coin lands heads, $\mathbb{E}$ is conducted, and ω is observed. Clearly the following two conditions are equivalent:

1. $Q^{\mathbb{E}}(H|\omega) = Q^{\mathbb{F}}(H|\omega)$ for any prior Q and hypothesis H.
2. $Q^{\mathbb{M}}(H|\langle \mathbb{E}, \omega \rangle) = Q^{\mathbb{M}}(H|\langle \mathbb{F}, \omega \rangle)$ for any prior Q and hypothesis H.

Notice the second condition requires only comparing posterior probabilities of outcomes of a single experiment, $\mathbb{M}$. Therefore, we use the second condition in our definition of *irrelevance* of stopping rules, as it generalizes to the qualitative setting in which one cannot directly compare posterior probabilities from different experiments.

Definition 9. Given two experiments $\mathbb{E}$ and $\mathbb{F}$ that differ only in stopping rule, we say knowledge of the stopping rule is $\mathbb{E}/\mathbb{F}$-*irrelevant* if, for any ω that can be obtained in both experiments, the outcomes $\langle \mathbb{E}, \omega \rangle$ and $\langle \mathbb{F}, \omega \rangle$ of $\mathbb{M}$ are posterior equivalent, where $\mathbb{M}$ is the mixed experiment in which a fair coin is flipped to decide whether to conduct $\mathbb{E}$ or $\mathbb{F}$.

Below, we argue that the choice between certain pairs of stopping rules is irrelevant in this sense. Our argument is a straightforward application of Corollary 1. We begin by identifying the stopping rules we will focus on and what it means for two experiments to "differ only" in stopping rule.

4.1 Non-Informative Stopping Rules

Often, the data-generating process in an experiment $\mathbb{E}$ is decomposable into two parts [Raiffa and Schlaifer, 1961, §2.3]. The first part

$$h^{\mathbb{E}}(\vec{x}; \theta_1) = f^{\mathbb{E}}(x_1; \theta_1) \cdot f^{\mathbb{E}}(x_2 | x_1; \theta_1) \cdots f^{\mathbb{E}}(x_n | x_1, \ldots x_{n-1}; \theta_1)$$

determines the chances that a measurement device takes a sequence of values $\vec{x} = \langle x_1, \ldots x_n \rangle$. Those probabilities are a function exclusively of the parameter of interest θ_1. The second component $\phi(n; \theta_1, \theta_2)$, which is called the *stopping rule*, determines the probability that an additional measurement is made at all, and its values might be affected by innumerable factors in addition to the value of the parameter of interest (e.g., budget, time, etc.); we can represent those factors by the nuisance parameter θ_2. The probability of stopping after n many points is

$$s^{\mathbb{E}}(n | \vec{x}; \theta_1, \theta_2) = \phi(x_1; \theta_1, \theta_2) \cdots \phi(n | x_1, \ldots, x_{n-1}; \theta_1, \theta_2) \cdot (1 - \phi(n+1 | \vec{x}; \theta_1, \theta_2)),$$

and so letting $\theta = \langle \theta_1, \theta_2 \rangle$, one can factor the likelihood function as follows:

$$P^{\mathbb{E}}_{\theta}(\vec{x}) = h^{\mathbb{E}}(\vec{x}; \theta_1) \cdot s^{\mathbb{E}}(n | \vec{x}; \theta_1, \theta_2)$$

A very special - but still controversial - case is when $s^{\mathbb{E}}$ depends on neither parameter and is either zero or or one, depending on $\vec{x}$. Such stopping rules form a special subset of what Raiffa and Schlaifer [1961] call "non-informative" stopping rules, and they can easily be seen to be irrelevant in the sense above. Why? Suppose two experiments $\mathbb{E}$ and $\mathbb{F}$ have the same data-generating mechanism $h(\vec{x}; \theta)$; this is what it means for the two experiments to "differ only" in stopping rule. Further, suppose both $s^{\mathbb{E}}$ and $s^{\mathbb{F}}$ depend on neither parameter, and assume that $s^{\mathbb{E}}(\vec{x})$ and $s^{\mathbb{E}}(\vec{y})$ are always either zero or or one. If a sample $\vec{x}$ can be obtained with positive probability in both experiments, then

$$P^{\mathbb{E}}_{\theta}(\vec{x}) = h(\vec{x}; \theta) \cdot s^{\mathbb{E}}(n | \vec{x}) = h(\vec{x}; \theta) = h(\vec{x}; \theta) \cdot s^{\mathbb{F}}(n | \vec{x}) = P^{\mathbb{F}}_{\theta}(\vec{x}).$$

Thus, the likelihood function of $\vec{x}$ is the *exactly the same* (not just proportional!) in both experiments, and so observing $\vec{x}$ in $\mathbb{E}$ is posterior equivalent to observing $\vec{x}$ in $\mathbb{F}$, as claimed. Notice that a likelihoodist who rejects a fully Bayesian viewpoint could also argue that the choice between such stopping rules is irrelevant because, by the likelihood principle, outcomes obtained in experiments that differ only by such stopping rules are evidentially equivalent.

Example 3: Suppose you are interested whether the fraction θ of people who survive three months of a new cancer treatment exceeds the survival rate of the conventional treatment, which is known to be 94%. Two experimental designs are considered: $\mathbb{E}$, which consists in treating 100 patients, and $\mathbb{F}$, which consists in treating patients until two deaths are recorded. In both experiments, it is possible to treat 100 patients and record precisely two deaths, the second of which is the 100th patient. Let $\vec{x}$ be a binary sequence of length 100 representing such a data set, and note that $P_\theta^{\mathbb{E}}(\vec{x}) = P_\theta^{\mathbb{F}}(\vec{x}) = \theta^2 \cdot (1-\theta)^{98}$ no matter the survival rate θ. Here, the stopping rules of the respective experiments are given by:

$$\phi^{\mathbb{E}}(n|x_1, \ldots x_{n-1}; \theta) = \begin{cases} 1 \text{ if } n < 100 \\ 0 \text{ otherwise} \end{cases}$$

and

$$\phi^{\mathbb{F}}(n|x_1, \ldots x_{n-1}; \theta) = \begin{cases} 1 \text{ if } \sum_{k \leq n-1} x_k < 2 \\ 0 \text{ otherwise} \end{cases}$$

By the arguments above, the choice between the two stopping rules are irrelevant from both a robust Bayesian and likelihoodist perspective. However, a quick calculation shows that under the null hypothesis – that the new drug is no more effective than the conventional treatment – the chance of two or fewer deaths in $\mathbb{E}$ is .0566, which is not significant at the .05 level. In contrast, in $\mathbb{F}$, the chance of treating 100 or more patients under the null is .014, which is significant at the .05 level. Advocates of classical testing sometimes conclude, therefore, that the latter experiment provides better evidence against the null than the former. This indicates a way in which LP is at odds with some classical testing procedures.

4.2 Qualitative Irrelevance of Stopping Rules

We now argue that, in our qualitative framework, the special class of stopping rules discussed in the previous section are irrelevant. Of course, it is unclear what the analog of "factoring" likelihood functions is in our qualitative framework, and so one might wonder what exactly a qualitative stopping rule is. That question can be avoided entirely. Notice that our argument above showed that, for the special class of stopping rules identified, if $\mathbb{E}$ and $\mathbb{F}$ "differ only" in stopping rule, then $P_\theta^{\mathbb{E}}(\omega) = P_\theta^{\mathbb{F}}(\omega)$ for all θ and all ω that can be obtained in both experiments. Then, it will also be the case that $P_\theta^{\mathbb{M}}(\langle \mathbb{E}, \omega \rangle) = P_\theta^{\mathbb{M}}(\langle \mathbb{F}, \omega \rangle)$ for all θ and all ω, where (1) $\mathbb{M}$ is the mixed experiment in which one flips a fair coin to decide whether to conduct $\mathbb{E}$ or $\mathbb{F}$ and (2) the outcomes $\langle \mathbb{E}, \omega \rangle$ and $\langle \mathbb{F}, \omega \rangle$ of $\mathbb{M}$ are as defined above.

Thus, whatever a "stopping rule" is in the qualitative setting, we can define two experiments to *differ only* in stopping rule (for the special class of stopping rules we've identified) if $\langle \mathbb{E}, \omega \rangle | \theta \equiv^{\mathbb{M}} \langle \mathbb{F}, \omega \rangle | \theta$ for all θ and any outcome ω obtainable in both experiments. By Corollary 1, if $\langle \mathbb{E}, \omega \rangle | \theta \equiv^{\mathbb{M}} \langle \mathbb{F}, \omega \rangle | \theta$ for all θ, then $\langle \mathbb{E}, \omega \rangle$ and $\langle \mathbb{F}, \omega \rangle$ are posterior and favoring equivalent. Thus, by the definition of "irrelevance" of stopping rule, knowledge of whether $\mathbb{E}$ or $\mathbb{F}$ was conducted is irrelevant. It should be noted that, just as in the quantitative case, a likelihoodist could also argue for the "irrelevance" of stopping rules by appealing to the qualitative analog of LP$^-$.

5. CONCLUSIONS AND FUTURE WORK

We have shown that, just as LL and LP characterize when all Bayesian reasoners agree how evidence should alter one's posterior probabilities, qualitative likelihoodist theses like QLL characterize agreement among agents whose beliefs satisfy weak coherence requirements. Our work should be extended in at least four ways.

First, one should characterize necessary conditions for qualitative posterior equivalence; we have proven only that our qualitative analog of LP^{-} is sufficient.

Second, our framework assumes the totality of the relation $\preceq$, but the most plausible generalization of Bayesian reasoning models agents as having degrees of beliefs that are not totally ordered but rather, are representable only by "imprecise" probabilities [Walley, 1991]. It is necessary to assess how many of our results continue to hold when $\preceq$ is only a partial order and to see what additional axioms might be appropriate in that case to secure the above results.

Third, we have assumed that the parameter space is finite. Assuming a qualitative countable additivity assumption [Krantz et al., 2006, p. 216], we conjecture, will allow us to extend those results to countable spaces, but it is unclear how to extend our framework to uncountable spaces.

Finally, Birnbaum [1962] began the exploration of the relationship between LP and a variety of other evidential "axioms", including the sufficiency, conditionality, and invariance principles. The relationship between those additional axioms should be explored in our qualitative framework.

REFERENCES

Shiri Alon and Ehud Lehrer. Subjective multi-prior probability: A representation of a partial likelihood relation. *Journal of Economic Theory*, 151:476–492, 2014.

James O. Berger. Robust Bayesian analysis: sensitivity to the prior. *Journal of statistical planning and inference*, 25(3):303–328, 1990. Publisher: Elsevier.

James Orvis Berger and Robert L. Wolpert. The likelihood principle. In *Ims Lecture Notes-Monograph*, volume 6. Institute of Mathematical Statistics, 1988.

David R. Bickel. The strength of statistical evidence for composite hypotheses: Inference to the best explanation. *Statistica Sinica*, pages 1147–1198, 2012.

Allan Birnbaum. On the foundations of statistical inference. *Journal of the American Statistical Association*, 57(298):269–306, 1962.

Anthony William Fairbank Edwards. *Likelihood*. CUP Archive, 1984.

Wade Edwards, Harold Lindman, and Leonard J. Savage. Bayesian Statistical Inference for Psychological Research. In Joseph B. Kadane, editor, *Robustness of Bayesian Analysis*, volume 4 of *Studies in Bayesian Econometrics*, pages 1–60. Elsevier Science Publishing Company Inc., Amsterdam: North Holland, 1984.

Michael J. Evans, Donald AS Fraser, and Georges Monette. On principles and arguments to likelihood. *Canadian Journal of Statistics*, 14(3):181–194, 1986.

D. A. S. Fraser, G. Monette, and K. W. Ng. Marginalization, likelihood and structured models. *Multivariate Analysis*, 6:209–217, 1984. Publisher: North-Holland Amsterdam.

Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. *Bayesian data analysis*. CRC press, 2013.

Ian Hacking. *Logic of statistical inference*. Cambridge University Press, 1965.

Joseph B. Kadane, editor. *Robustness of Bayesian Analysis*, volume 4 of *Studies in Bayesian Econometrics*. Elsevier Science Publishing Company Inc., Amsterdam: North Holland, 1984.

David H. Krantz, R. Duncan Luce, Patrick Suppes, and Amos Tversky. *Foundations of Measurement Volume II: Geometrical, Threshold, and Probabilistic Representations*. Dover Publications, Mineola, N.Y, December 2006. ISBN 978-0-486-45315-6.

Howard Raiffa and Robert Schlaifer. *Applied statistical decision theory*. Wiley Classics Library.

Wiley-Interscience, 1961. Publisher: Division of Research, Graduate School of Business Adminitration, Harvard ....

Richard Royall. *Statistical evidence: a likelihood paradigm*, volume 71. Chapman & Hall/CRC, 1997.

Leonard J. Savage, George Barnard, Jerome Cornfield, Irwin Bross, I. J. Good, D. V. Lindley, C. W. Clunies-Ross, John W. Pratt, Howard Levene, and Thomas Goldman. On the foundations of statistical inference: Discussion. *Journal of the American Statistical Association*, 57 (298):307–326, 1962. Publisher: JSTOR.

Elliott Sober. *Evidence and evolution: The logic behind the science*. Cambridge University Press, 2008.

Peter Walley. *Statistical reasoning with imprecise probabilities*. Number 42 in Monographs on Statistics and Applied Probability. Chapman & Hall, New York, 1991.

Sérgio Wechsler, Carlos A. de B. Pereira, Paulo C. Marques F., Marcelo de Souza Lauretto, Carlos Alberto de Bragança Pereira, and Julio Michael Stern. Birnbaum's Theorem Redux. *AIP Conference Proceedings*, 1073(1):96–100, November 2008. ISSN 0094-243X. . URL https://aip.scitation.org/doi/abs/10.1063/1.3039028. Publisher: American Institute of Physics.